

(11)Publication number : 03-286259  
(43)Date of publication of application : 17.12.1991

G06F 15/40

(71)Applicant : SHIMADZU CORP

(72)Inventor : NISHI NOBUHIKO  
NAKAI KOICHI

(57)Abstract:

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A) 平3-286259

⑬ Int. Cl. 5

識別記号

庁内整理番号

⑭ 公開 平成3年(1991)12月17日

G 06 F 15/40

5 1 0 N

7056-SL

審査請求 未請求 請求項の数 1 (全6頁)

⑮ 発明の名称 データベースの統制語選出装置

⑯ 特 願 平2-87074

⑰ 出 願 平2(1990)3月30日

⑱ 発 明 者 西 伸 彦 京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所三条工場内

⑲ 発 明 者 中 井 光 一 京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所三条工場内

⑳ 出 願 人 株式会社島津製作所 京都府京都市中京区西ノ京桑原町1番地

㉑ 代 理 人 弁理士 杉 谷 勉

明 細 書

1. 発明の名称

データベースの統制語選出装置

2. 特許請求の範囲

(1) 多数のデータを格納するデータベースと、前記データベース検索に用いられる統制語および統制語と同義である複数個の同義語を記憶する同義語辞書と、ユーザが入力(使用)した同義語を前記同義語辞書を用いて統制語に変換し、その統制語でデータベース中から所望のデータを検索するデータベース検索手段と、前記各同義語の使用頻度に応じた使用係数値を記憶する使用係数記憶部と、前記同義語が使用される毎に予め設定された変更量を用いて、使用された同義語およびその他の同義語の使用係数値を更新する使用係数更新手段と、予め設定された時間毎に全ての同義語の使用係数値が初期値に近づくように、予め設定された時間の変更量を用いて使用係数値の時間的変更を行う時間的使用係数更新手段と、ユーザから統制語の変更要求が出されると、前記各手段によ

って、更新された使用係数値のうち最も高い係数値を有する同義語を新たな統制語として更新する統制語更新手段とを備えたことを特徴とするデータベースの統制語選出装置。

3. 発明の詳細な説明

A. 産業上の利用分野

この発明は、統制語を用いてデータベース内のデータを検索するデータベースシステムで利用され、特に、その統制語を選出するため技術に関する。

B. 従来技術

従来、文献データのように、1つの言葉に対する呼称が複数個存在するようなデータをデータベース中から検索する際、同義語辞書を用いてユーザから入力されたキーワードを統制語(control term)と表記され、以下C.T.とも略称する)に変換し、このC.T.でデータベース内を検索して所望のデータを抽出するデータベースシステムが知られている。

その同義語辞書とは、統制語に対する複数個の

同義語(統制語と同義の言葉ではあるが、呼称が異なったもの)を記憶したもので、同義語辞書を用いてデータベース内を検索した結果、出力されるデータの見出しには、その統制語が使われる。

したがって、統制語としては、現在最も多く使用されている言葉が適当である。逆に、過去には多く使用されたが、現在は死語となっているような言葉を統制語として用いると、検索の結果、出力されるデータの見出しを見たときにユーザ混乱を恐る恐れがあり、望ましいことではない。

このため、同義語辞書中、最も多く使用されている同義語を新たな統制語として置き換える作業をある一定期間毎に行う必要があり、従来、以下のようにして行われている。

ある設定時間内における各同義語の使用頻度を計測し、計測結果である同義語の使用頻度をユーザに提示する。ユーザは提示された各同義語とそれらの使用頻度から、最も高い使用頻度である同義語を新たな統制語として指定する。ユーザからの指定があった同義語を新たな統制語として更新

する。

C. 発明が解決しようとする課題

しかしながら、上述した従来技術には次のような問題点がある。

即ち、統制語の書き換えを行う際に、ある設定期間を設け、その期間内における同義語の使用回数を計測する必要があるため、ユーザから統制語の変更要求が出されたときに、迅速に同義語の使用頻度をユーザに提示することができず、統制語の更新作業の効率低下を招くという問題点がある。

この発明は、このような事情に臨みてなされたものであって、迅速に同義語の使用頻度をユーザに提示することができるデータベースの統制語選出装置を提供することを目的としている。

D. 課題を解決するための手段

この発明は、上記目的を達成するために次のような構成を備えている。

即ち、この発明に係るデータベースの統制語選出装置は、多数のデータを格納するデータベースと、前記データベース検索に用いられる統制語お

よび統制語と同義である複数箇の同義語を記憶する同義語辞書と、ユーザが入力(使用)した同義語を前記同義語辞書を用いて統制語に変換し、その統制語でデータベース中から所望のデータを検索するデータベース検索手段と、前記各同義語の使用頻度に応じた使用係数値を記憶する使用係数記憶部と、前記同義語が使用される毎に予め設定された変更値を用いて、使用された同義語およびその他の同義語の使用係数値を更新する使用係数更新手段と、予め設定された時間毎に全ての同義語の使用係数値が初期値に近づくように、予め設定された時間的変更値を用いて使用係数値の時間的変更を行う時間的使用係数更新手段と、ユーザから統制語の変更要求が出されると、前記各手段によって、更新された使用係数値のうち最も高い係数値を有する同義語を新たな統制語として更新する統制語更新手段とを備えたことを特徴としている。

E. 作 用

この発明のデータベースの統制語選出装置を用

いた統制語の選出は、次のようにして行われる。

まず、ユーザがデータベース検索用のキーワードとして、統制語と同義である同義語を入力すると、データベース検索手段は、同義語辞書を用いて入力された同義語を統制語に変換し、その統制語でデータベース中から所望のデータを検索する。このようなデータベースの検索が行なわれる度に、使用係数更新手段は、使用係数記憶部内からユーザが入力(使用)した同義語およびその他の同義語の使用係数値を読み出し、予め設定された変更値を用いて、それらの使用係数値を更新していく。これにより、ユーザの使用頻度に応じて同義語の使用係数値が更新される。

一方、時間的使用係数更新手段は、予め設定された時間になれば、前記使用係数記憶部内の全ての使用係数値を読み出し、予め設定された時間的変更値を用いて、全ての使用係数値が初期値に近づくように使用係数値の更新を行う。このように、過去に累積された使用係数値は、時間の経過とともに初期値に近づくように更新されるので、使用

係数記憶部内に記憶される使用係数値は現時点における使用頻度を近似的に表した値となる。

統制部更新手段は、ユーザから統制部変更の要求が出されると、最も高い使用係数値をもつ同義語を新たな統制語として更新する。

#### F. 実施例

以下、この発明の実施例を図面に基づいて説明する。

第1図は、この発明のデータベースの統制部選出装置の概略構成を示したブロック図である。

図中、符号1は予め多数のデータを格納するデータベースであり、このデータベース1には、同義語辞書2が備えられている。同義語辞書2は従来技術での説明にもあるように、データベース1内を検索するときに用いる統制語と、その統制語に対する複数個の同義語を記憶したものである。符号3は、キーボードやディスプレイなどを備えた入力部である。この入力部3はユーザからの検索用キーワード(統制語に対する自然語であり、以下 free term=PTとも略称する)を入力した

り、検索結果を表示するなどの用途に使用される。

4は、同義語辞書2を参照して入力部3から入力されたPTを統制語(CT)に変換し、このCTでデータベース1内のデータ検索を行い、検索結果を入力部3のディスプレイに表示するデータ検索手段としてのデータ検索部である。これらの構成部品で、データベース1内のデータ検索が行われる。

ここで、データベース1内に格納されているデータの一例として、工学系の文献データを扱い、ユーザから「モノクロメータ」に関する文献データの検索要求があったとする。しかし、ユーザからは、「モノクロメータ」ではなく、「モノクロ」が検索用のキーワードとして入力されたとする。

データ検索部4は、入力されたキーワード「モノクロ」を用いて同義語辞書2内の検索を行い、「モノクロ」のCTとして登録されている言葉を抽出する。いま、「モノクロメータ」が「モノクロ」の統制語として登録されていると、この「モ

ノクロメータ」を抽出することで、「モノクロ」を「モノクロメータ」に変換し、この「モノクロメータ」でデータベース1内の検索を行う。検索結果として、「モノクロメータ」に関する文献データを入力部3のディスプレイに表示する。その一方、データ検索部4は検索で使用したCT「モノクロメータ」と、ユーザから入力されたPT「モノクロ」とを第1図示の使用係数変更部5に送出する。

使用係数変更部5は、与えられた「モノクロメータ」に対する全ての同義語と、各同義語の使用係数値を、使用係数記憶部としての同義語使用係数記憶部7から読み出す。

同義語使用係数記憶部7内に格納されているデータの構造例を第2図に示す。

この図に示すように、同義語使用係数記憶部5内には、CT「モノクロメータ」に対する全ての同義語「分光器」、「モノクロメータ」(これが現時点における統制語になっている)、「モノクロ」、「光学解析器」と、これらの使用係数値

がデータとして記憶されている。使用係数値は、同義語の使用頻度を表す数値であり、この例では、「1」を最大値、「0」を最小値に設定している。また、各同義語の使用係数としては、予め初期値が設定されており、この例では初期値として「0.5」を設定している。

使用係数変更部5は、このような1つのCTに対する全てのPTと使用係数値を読み出した後、読み出した使用係数値に対応して記憶されている使用係数変更値(以下、単に変更値と略称する)を使用係数変更値記憶部6から読み出す。

この使用係数変更値メモリ7に記憶されている変更値の一例をグラフ形式にして第3図に示す。この図に示すように、変更値は変更後の使用係数値が、設定されている最大値「1」よりも大きく、また、最小値「0」よりも小さくならないように、使用係数値が「1」と「0」のときに「0」の値をとるように設定されている。このため、中心値である使用係数「0.5」のときに最大値をとるような値に設定されている。この最大値の一例とし

ては「0.01」が設定されている。

いま、同義語使用係数記憶部7から読み出された各同義語の使用係数は、全て初期値である「0.5」に設定されているので、使用係数変更部5は使用係数「0.5」を読み出しアドレスとして、使用係数変更記憶部6に与え、使用係数値「0.5」に対応している変更値「0.01」の読み出しを行う。

変更値の読み出し後、使用係数変更部5は、使用係数変更記憶部6から読み出した変更値「0.01」でもって、各同義語の使用係数値を変更する。この変更は、データ検索部4から送られてきたF-T「モノクロ」(ユーザから入力されたキーワード)の使用係数値に変更値「0.01」を加算し、残りの同義語の使用係数値から変更値「0.01」を減算することで行われる。

変更の結果、各同義語の使用係数値は以下のようになる。

「分光器」→0.5 - 0.01 = 0.49

「モノクロメータ」→0.5 - 0.01 = 0.49

次に、その時間的な使用係数値の変更について説明する。

この変更を実施する構成部品は、第1図示の全使用係数変更部8と時間的変更記憶部9、タイマ図10であり、これらの部品がこの発明の時間的使用係数更新手段に相当している。

タイマ図10は、全使用係数変更部8に対して常時、時刻データを送出しており、全使用係数変更部8はタイマ図10から送られた時刻データが設定された時間になれば、使用係数値の時間的使用係数更新を実行する。その設定時間はユーザによって、任意に設定されるもので、例えば「一日」を設定時間とした場合には、一日毎に全ての使用係数値の変更が以下のようにして行われる。

全使用係数変更部8は、同義語使用係数記憶部7内に記憶されている同義語と、その使用係数値を記憶順に次々と読み出し、読み出した使用係数値に対応する時間的使用係数更新手段を時間的使用係数記憶部9から読み出す。

時間的使用係数記憶部9内に記憶されているデー

「モノクロ」→0.5 + 0.01 = 0.51

「光学分解器」→0.5 - 0.01 = 0.49

このような演算を行った後、使用係数変更部5はその演算結果で、同義語使用係数記憶部7内の使用係数値を更新する。

データベース1内のデータ検索が行われるたびに上記のようにして、使用係数値の変更処理が実行され、ユーザから入力(使用)されたF-Tの使用係数値は増加し、使用されなかったF-Tの使用係数値は減少する。上記使用係数変更部5と使用係数変更記憶部6とがこの発明の使用係数更新手段に相当している。

このように、ユーザが入力キーワードとして用いた同義語の使用係数値を上げ、他の同義語の使用係数を下げるという処理だけでは、「過去に頻繁に使用されたが、現在は死語になっている言葉」かどうかの判断基準にはなり得ない。そこで、過去の使用頻度の影響を打ち消すために、ある設定時間ごとに、全ての使用係数値を時間的に変更する必要がある。

タ例をグラフ形式にして、第4図に示す。この図に示すように、時間的使用係数記憶部9内に記憶されているデータは、過去に算出された使用係数の影響を打ち消すために最大使用係数値「1」には、最小の時間的使用係数(この例ではマイナス「0.01」)が設定され、最小使用係数値である「0」に対しては、最大の時間的使用係数であるプラス「0.01」が設定されている。したがって、使用係数値の中心値である「0.5」に対しては、時間的使用係数「0」が設定されている。

全使用係数変更部8は、同義語使用係数記憶部6から読み出した使用係数値を、時間的使用係数値読み出しアドレスとして時間的使用係数メモリ9に与え、使用係数値に対応して記憶されている変更値を読み出す。例えば、前述の例である「分光器」の使用係数値は、「0.49」であるため、「0.49」に対応して記憶されている時間的使用係数、この例では「+0.001」が読み出され、「モノクロ」の使用係数値は「0.51」であるため、「0.51」に対応して記憶されている時間的使用係数、この例では

「-0.001」が読み出される。

変更鍵を踏み出した後、全使用係数変更部8は、使用係数値に変更鍵を加算する。

この結果、上記例の使用係数値は以下のように変更される。

「分光器」→ $0.49+0.001=0.491$

「モノクロ」→ $0.51-0.001=0.509$

;

この使用係数値の時間的変更を、同義語使用係数記憶部7内の全ての同義語の使用係数値に対して行った後、変更した使用係数値で同義語使用係数記憶部7内の使用係数値を更新する。

このようにして、同義語使用係数記憶部7内に記憶されている使用係数値は、単に使用頻度に応じた値ではなく、過去の影響が打ち消されることにより、現時点で最も多く使用されていることを示した値となる。

ユーザからC Tの変更要求指示が、入力部3のキーボードを介して与えられると、第1図示のC T変更部11は、同義語使用係数記憶部7内の同義

語とその使用係数値を入力部3のディスプレイに表示する。ユーザは表示された使用係数値をみて、最も高い係数値をもつ同義語を新たなC Tとするように再びキーボードを介して指示する。C T変更部11は指示のあった同義語を新たなC Tとして、登録し直す。このC T変更部11はこの発明の統制語更新手段に相当している。

このように、ユーザからC T変更の要求が出されたときには、迅速に各同義語の使用係数値(現時点における使用頻度を示した値)をユーザに提示することができる。

なお、上述した実施例では、ユーザから統制語の更新指示が出されると、C T変更部11は同義語およびその使用係数値を入力部3のディスプレイに表示して、ユーザに知らせようとしたが、これはディスプレイに表示することなく、ユーザからの更新指示が出されると、C T変更部11が自動的に各同義語の使用係数値を読み込み、最も高い使用係数値を有する同義語を新たな統制語として更新するようにしてもよい。

#### G. 発明の効果

以上の説明から明らかなように、この発明に係るデータベースの統制語選出装置によれば、データベース検索を行う際に、ユーザが入力(使用)した同義語およびその他の同義語の使用係数値を更新し、さらに、最定時間前に前記使用係数値が初期値に近づくように更新するので、使用係数値記憶部内には、常に近似的に現時点における使用頻度を表した使用係数値が記憶される。したがって、ユーザから統制語の変更要求が出されると、即座に、現時点における使用係数値をもって、統制語の更新を行うことができるから、従来の統制語更新処理に比べ、大幅にその処理効率の向上を図ることができる。

#### 4. 図面の簡単な説明

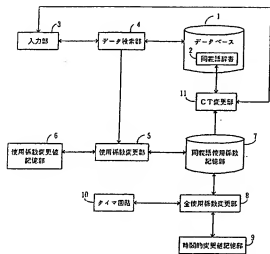
第1図ないし第4図は、この発明の一実施例に係る、第1図はデータベースの統制語選出装置の概略構成を示したブロック図、第2図は同義語使用係数記憶部内のデータ構造の一例を示した図、第3図は使用係数変更記憶部内に記憶されてい

るデータの一例をグラフ形式にして示した図、第4図は時間的変更記憶部内に記憶されているデータの一例をグラフ形式にして示した図である。

- 1…データベース 2…同義語辞書
- 4…データ検索部 5…使用係数変更部
- 6…使用係数変更記憶部
- 7…同義語使用係数記憶部
- 8…全使用係数変更部
- 9…時間的変更記憶部 10…タイム図路
- 11…C T変更部

特許出願人 株式会社 鳥津製作所  
代理人 弁理士 杉谷 勉

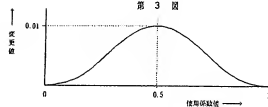
第 1 図



第 2 図

C.T.	S.T.	使用部記憶
モノクローム	カラー	0.5
モノクローム	モノクローム	0.5
モノクローム	モノクローム	0.5
モノクローム	モノクローム	0.5
モノクローム	モノクローム	0.5
モノクローム	モノクローム	0.5
モノクローム	モノクローム	0.5
モノクローム	モノクローム	0.5

第 3 図



第 4 図

